

**IDENTIFIKASI ARTI TANGIS BAYI
VERSI DUNSTAN *BABY LANGUAGE* MENGGUNAKAN
JARAK TERPENDEK DARI JARAK MAHALANOBIS
(Infant Cries Identification of Dunstan Baby Language Version using the
Shortest Distance of Mahalanobis)**

¹Medhanita Dewi Renanti

¹Program Diploma Institut Pertanian Bogor, medha_nuha@yahoo.com

Diterima 7 Oktober 2013/ Disetujui 21 Oktober 2013

ABSTRACT

New born babies have the ability to express their basic needs through sounds. A system to understand the meaning of crying infants of aged 0-3 months is called Dunstan Baby Language (DBL), which was introduced in 2006. This research aimed to perform the modeling of codebook method with k-means clustering technique as feature matching, and Mel Frequency Cepstrum Coefficients (MFCC) as feature extraction to identify the infant cries. The infant cries identification of Dunstan Baby Language version used the shortest distance of mahalanobis. The treatment in this research was the combination of frame length: 25 ms, 40 ms and 60 ms, frame overlap of 0%, 40%, and 60%, and the number of codewords (number of clusters) of 1 to 29. The best accuracy in recognizing five types of crying Infants using mahalanobis distance can be achieved up to 83% when the frame length = 275, the overlap frame = 0.25, and the k = 17. Sound 'heh' was the most familiar, whereas sound 'owh' was always misunderstood and generally known as 'neh' and 'eairh'.

Keywords: Codebook, Dunstan baby language, Mahalanobis Distance, MFCC

PENDAHULUAN

Dunstan (2006) menyatakan bahwa setiap bayi lahir mempunyai kemampuan untuk mengekspresikan kebutuhan dasar mereka melalui suara. Penelitian yang dilakukan oleh Priscilla Dunstan dari tahun 1998 sampai 2006 telah berhasil menemukan bahasa yang dimiliki oleh semua bayi di seluruh dunia usia 0-3 bulan. Bahasa tersebut dikenal dengan sebutan *Dunstan Baby Language*-DBL.

Terdapat lima bahasa bayi versi DBL yaitu: "neh" berarti lapar, "owh" berarti lelah yang mengindikasikan bayi sudah mulai mengantuk, "eh" berarti ingin bersendawa, "eairh" berarti nyeri (ada angin) di perut, dan "heh" berarti tidak nyaman (bisa karena popoknya basah, udara terlalu panas atau dingin, atau hal lainnya). Perintis DBL di Indonesia adalah dokter Adhiatma Gunawan yang menyebutkan bahwa seorang bayi mempunyai refleks primitif yang dimiliki sejak dilahirkan. Refleksi ini bersifat universal dan lambat laun akan menghilang seiring dengan berkembangnya kemampuan untuk beradaptasi. DBL berlaku pada bayi hingga usia tiga bulan karena setelah usia tersebut bayi akan mengembangkan kemampuan berkomunikasi sendiri dengan bantuan orang tua dan lingkungan. Penelitian membuktikan bahwa 90% dari ibu-ibu di seluruh dunia yang mengikuti DBL merasa puas dan terbantu dengan sistem tersebut dan merekomendasikan pada yang lain. Sementara lebih dari 70% orang tua merasa lebih percaya diri dalam mengasuh bayinya dan *by the end of the day*, tingkat stres mereka berkurang signifikan (Gunawan 2011).

Kepakaran untuk menentukan arti tangis bayi versi DBL masih agak jarang sehingga informasi arti tangis bayi ini belum mudah didapatkan oleh para orang tua. Saat ini sistem untuk mentransfer pengetahuan mengenai DBL dengan cara mengikuti pelatihan atau seminar yang diadakan oleh pengelola atau dengan cara mempelajari sendiri materi jenis tangis bayi versi DBL dalam bentuk *optical discs* atau dapat juga mengunduh materi DBL di internet. Pengguna sistem DBL khususnya di Indonesia akan lebih yakin dengan kesimpulan yang mereka buat jika ada suatu *software* yang secara otomatis dapat menghasilkan arti tangisan bayinya. Hal ini dapat menguatkan kesimpulan mereka, karena jika hanya mengikuti pelatihan kilat atau seminar, masih ada beberapa peserta yang belum paham bagaimana mengenali maksud tangisan bayi tersebut atau belum tepat memaknai tangisan bayi. Selain itu nantinya *software* ini juga bisa bermanfaat bagi orang tua yang belum mengikuti pelatihan atau seminar DBL sehingga para orang tua dapat memahami bahasa/tangis bayi.

Penelitian mengenai tangis bayi telah dilakukan oleh para peneliti, antara lain: klasifikasi tangis bayi normal dan abnormal (menderita gangguan hipoksia-tubuh kekurangan oksigen) menggunakan *neural network* menghasilkan akurasi 85% (Poel & Ekkel 2006). Klasifikasi bayi sehat dan bayi yang mengalami sakit seperti kerusakan otak, bibir sumbing, hidrosefalus, dan sindrom kematian bayi mendadak menggunakan metode klasifikasi *Hidden Markov Model*-HMM menghasilkan akurasi 91% (Lederman *et al.* 2008). Penelitian lainnya adalah klasifikasi tiga jenis tangis yaitu bayi normal, bayi tuli, dan bayi yang menderita asfiksia (tidak dapat bernafas secara spontan dan teratur) pada usia satu hari sampai sembilan bulan menggunakan *neural network* menghasilkan akurasi 86% (Reyes-Galaviz dan Reyes-Garcia 2004).

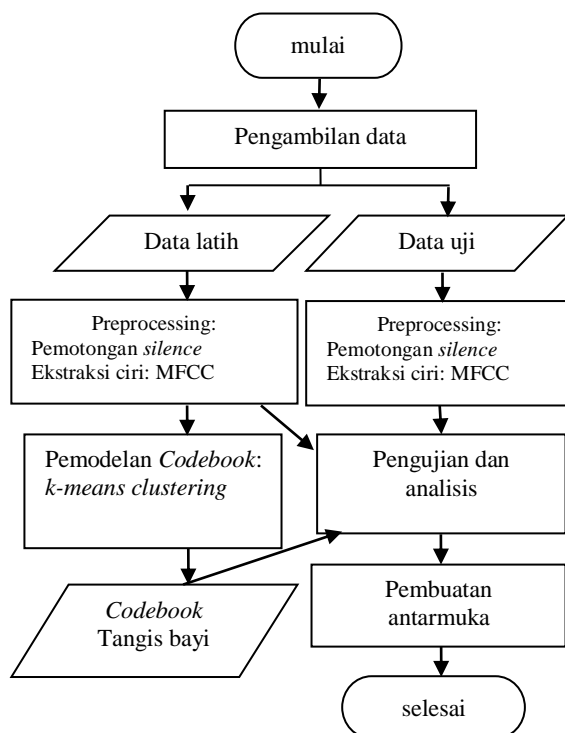
Dari beberapa hasil penelitian tersebut dapat disimpulkan bahwa tangisan bayi berarti sinyal suara tangis bayi dapat digunakan untuk mendeteksi status kesehatan bayi. Hal ini sejalan dengan penelitian yang dilakukan oleh Priscilla Dunstan yang menyatakan bahwa tangis bayi sebagai alat komunikasi untuk memenuhi kebutuhan atau keinginannya dan tangis bayi mempunyai tujuan untuk mengungkapkan sesuatu.

Penelitian klasifikasi tangis bayi sebelumnya telah menggunakan *neural network* atau HMM sebagai pengenalan polanya. Penelitian untuk identifikasi jenis bayi versi DBL ini menggunakan *codebook* untuk pengenalan

polanya dan *Mel-Frequency Cepstrum Coefficients* (MFCC) untuk ekstraksi ciri. Pemilihan metode ini didasari beberapa penelitian diantaranya: penelitian Lee *et al.* (2006) melakukan pengenalan suara burung menggunakan MFCC dan *Vector Quantization* (VQ) *codebook* dan berhasil mencapai akurasi 87%. Selain itu *speaker recognition system* juga berhasil dibuat menggunakan MFCC dan VQ (Kumar 2011). Penelitian serupa dilakukan oleh Singh dan Rajan (2011) yang berhasil mencapai akurasi 98,57% dengan melakukan penelitian *speaker recognition* menggunakan VQ dan MFCC.

METODOLOGI

Metode penelitian ini terdiri atas beberapa tahapan proses yaitu pengambilan data, *preprocessing*, pemodelan *codebook* tangis bayi, pengujian dan analisis, serta pembuatan antarmuka. Metodologi penelitian proses identifikasi jenis tangis bayi ditampilkan pada Gambar 1.



Gambar 1 Metodologi penelitian identifikasi arti tangis bayi

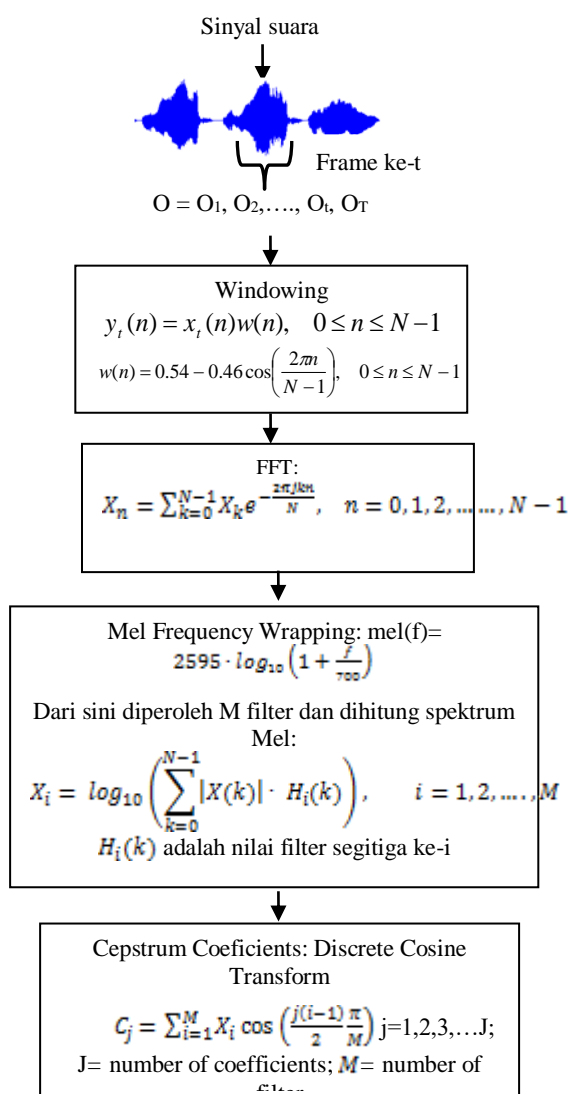
Pengumpulan Data

Data yang digunakan untuk penelitian ini adalah diambil dari video *Dunstan Baby Language* yang sudah dilakukan pengolahan data. Data terbagi menjadi dua yaitu data latih dan data uji. Terdapat 140 data latih yang masing-masing mewakili 28 tangis bayi lapar, 28 tangis bayi mengantuk, 28 tangis bayi ingin bersendawa, 28 tangis bayi mengalami nyeri (ada angin) di perut, dan 28 tangis bayi tidak nyaman (bisa karena popoknya basah/udara yang terlalu panas/dingin atau hal lainnya). Data uji sebanyak 35, masing-masing 7 tangis bayi untuk setiap jenis tangis bayi. *Sampling rate* yang digunakan pada penelitian ini adalah 11000Hz.

Preprocessing

Pada tahap *preprocessing* ini dilakukan pemotongan *silence* dan ekstraksi ciri menggunakan metode *Mel Frequency Cepstrum Coefficients* (MFCC). Pemotongan *silence* ini dilakukan di awal dan di akhir sinyal suara tangis bayi.

Ekstraksi ciri merupakan proses untuk menentukan satu nilai atau vektor yang dapat digunakan sebagai penciri objek atau individu. Ciri yang biasa digunakan adalah koefisien *cepstral* dari sebuah *frame*. MFCC merupakan ekstraksi ciri yang menghitung koefisien *cepstral* dengan mempertimbangkan pendengaran manusia (Buono 2009). Diagram alur dari MFCC dapat dilihat pada Gambar 2.



Gambar 2 Diagram Alur MFCC

Pemodelan *Codebook* Tangis Bayi

Pemodelan *codebook* ini berasal dari setiap data latih yang nantinya digunakan sebagai representasi sinyal tangis bayi yang akan dikenali. *Codebook* yang akan dibuat yaitu *codebook* dari setiap data tangis bayi. *Codebook* dari *cluster* yang dibuat berasal dari hasil *clustering* semua data tangis bayi dengan menggunakan *k-means clustering*.

Vector Quantization (VQ) merupakan proses memetakan vektor-vektor dari ruang vektor yang besar menjadi ruang vektor dengan jumlah terbatas. Setiap daerah disebut *cluster* dan dapat direpresentasikan oleh pusatnya yang disebut *codeword*. Kumpulan *codeword* disebut *codebook* (Do 1994). Pemodelan *speaker* menggunakan pendekatan berbasis VQ dibentuk oleh *clustering* dari fitur *speaker* pada K yang tidak

overlapping. Setiap *cluster* direpresentasikan oleh *code vector* c_i yang disebut centroid. Hasil himpunan *code vector* ini disebut *codebook*. *Codebook* ini berfungsi sebagai model pembicara (Linde *et al.* 1980).

K-means clustering merupakan algoritma *clustering* eksklusif. Setiap objek ditetapkan secara tepat ke dalam salah satu dari sekumpulan *cluster*. Metode *clustering* ini dimulai dengan menentukan banyaknya *cluster* yang akan dibentuk (disebut dengan nilai k). Nilai k umumnya nilai integer terkecil seperti 2, 3, 4, atau 5. Pengukuran kualitas satu *cluster* dapat menggunakan nilai dari fungsi objektif yang diambil dari jumlah kuadrat jarak setiap titik dari *centroid* yang telah ditentukan (*euclidean distance*). Pilih poin k yang umumnya sesuai dengan lokasi k dari suatu objek. Poin k ini nantinya dijadikan sebagai *centroid* dari k *cluster*. Metode ini dapat bekerja maksimal jika pemilihan poin awal k yang berjauhan. Selanjutnya dipilih setiap poin, satu per satu dari suatu *cluster* yang mempunyai *centroid* terdekat. Semua objek mempunyai k *cluster* berdasarkan k *initial centroid* tetapi '*centroid*' ini tidak akan menjadi *centroid* sejati suatu *cluster*. Langkah selanjutnya dihitung ulang *centroid* dari *cluster* dan dilakukan langkah-langkah sebelumnya untuk menempatkan setiap objek ke *cluster* dengan *centroid* terdekat. Algoritma *k-means clustering* sebagai berikut (Bramer 2007):

1. Pilih nilai k
2. Pilih objek k secara acak. Objek ini akan menjadi k initial centroid
3. Tetapkan setiap objek dari suatu cluster yang terdekat dengan centroid
4. Hitung ulang centroid dari k cluster
5. Ulangi langkah 3 dan 4 sampai perubahan nilai centroid pada iterasi satu dengan lainnya tetap (centroid tidak bergerak)

Tahap pembuatan *codebook* untuk seorang pembicara sebagai berikut (Buono 2009):

1. Untuk setiap pengucapan (ada n pengucapan sebagai data *training*), dilakukan ekstraksi ciri menggunakan MFCC pada setiap *frame* dengan panjang dan *overlap* tertentu.
2. Semua *frame* dari n pengucapan digabungkan menjadi satu set dan dilakukan *unsupervised clustering* menggunakan teknik *k-means clustering* dengan memilih banyaknya *cluster* sesuai dengan jumlah *codeword* yang diinginkan.

Pengujian dan Analisis

Tahapan pengujian berarti melakukan pengujian untuk identifikasi jenis tangis bayi. Alur proses untuk tahap identifikasi/pengenalan adalah (Buono 2009):

1. Untuk setiap ucapan baru yang masuk ke sistem dibaca *frame* demi *frame*, (misalkan jumlah *frame* yang diperoleh adalah T), dan dilakukan ekstraksi ciri menggunakan MFCC.
2. Hitung jarak sinyal *input* ucapan ini ke *codebook* setiap pembicara yang ada dalam sistem.
3. Keputusan: menetapkan label pada *input* suara sesuai dengan pembicara dengan jarak *codebook* terkecil.

Jarak *input* ucapan dengan *codebook* dirumuskan sebagai berikut (Buono 2009):

1. Untuk setiap *frame* dari *input* ucapan yang masuk, hitung jarak ke setiap *codeword* dan dipilih *codeword* dengan jarak minimum.
2. Jarak antara *input* ucapan dengan *codebook* adalah jumlah dari jarak minimum tersebut (persamaan 1):

$$\text{jarak}(\text{input}, \text{codebook}) = \sum_{t=1}^T \min_{k} [d(\text{frame}_t, \text{codeword}_k)] \quad (1)$$

Penelitian ini menggunakan jarak mahalanobis untuk identifikasi arti tangis bayi. Jarak mahalanobis didefinisikan pada persamaan 2 (Gomathy *et al.* 2012).

$$D = \sqrt{(x_s - y_s) * C^{-1} * (x_s - y_s)^T} \quad (2)$$

C adalah *covariance matrix*, $x = \{x_1, x_2, \dots, x_s\}^T$ dan $y = \{y_1, y_2, \dots, y_s\}^T$

Pengujian dilakukan menggunakan data uji sebanyak 35 dan data latih yang berjumlah 140. Tahap analisis dilakukan berdasarkan hasil yang didapatkan pada tahap pengujian. Analisis dilakukan berdasarkan hasil dari kombinasi faktor dan level berikut:

- panjang *frame* : 25 ms/panjang frame 275, 40 ms/ panjang frame 440, 60 ms/panjang frame 660.
- *overlap frame* : 0%, 25%, 40%.
- banyaknya *codeword*/jumlah *cluster*: 1 sampai dengan 18. Untuk panjang frame = 275 dan *overlap frame* = 0, jumlah k adalah 1-29 *cluster*.

Masing-masing kombinasi faktor dan level tersebut akan dihitung nilai akurasi menggunakan persamaan 3.

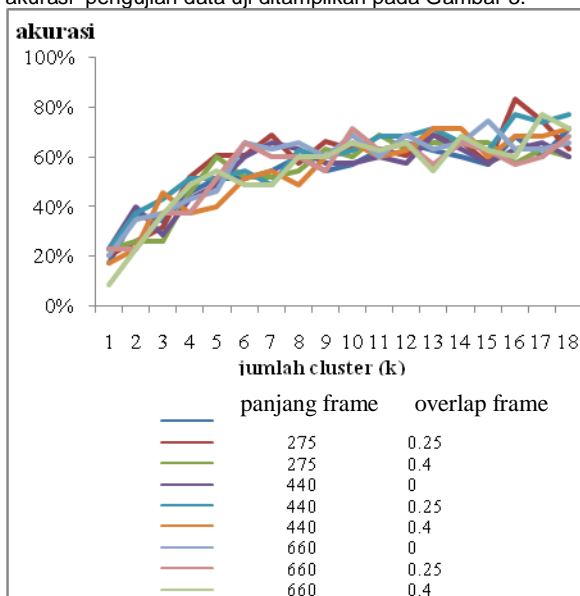
$$\text{akurasi} = \frac{\text{jumlah data pengujian dengan identifikasi benar}}{\text{jumlah data pengujian}} * 100\%$$

Pembuatan Antarmuka

Pembuatan antarmuka identifikasi jenis tangis bayi dibuat berdasarkan data latih yang menghasilkan akurasi tertinggi.

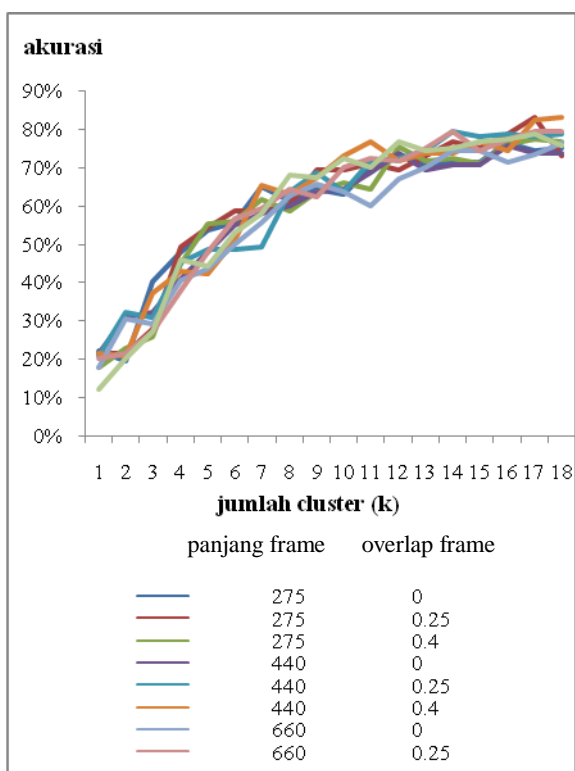
HASIL DAN PEMBAHASAN

Penelitian ini dibuat menggunakan *software* Matlab R2010b version 7.11.0.584. Hasil perbandingan akurasi pengujian data uji ditampilkan pada Gambar 3.



Gambar 3 Grafik perbandingan pengujian data uji

Grafik perbandingan akurasi menggunakan pengujian data latih ditampilkan pada Gambar 4.



Gambar 4 Grafik perbandingan pengujian data latih

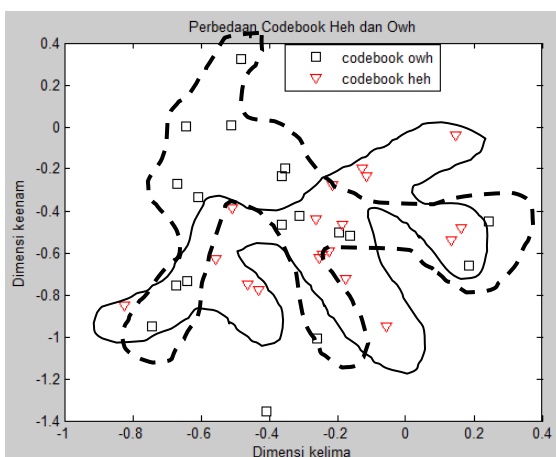
Akurasi menggunakan pengujian data uji berkisar antara 9% sampai 83%. Nilai akurasi 83% didapat ketika panjang frame = 275, overlap frame 0.25, dan k=16. Sedangkan akurasi pengujian menggunakan data latih berkisar antara 12% sampai 83%. Nilai akurasi 83% diperoleh ketika panjang frame = 275, overlap frame 0.25, dan k=17 serta panjang frame = 440, overlap frame 0.4, dan k=18. Dari beberapa model yang menghasilkan akurasi tertinggi tersebut, model yang akan digunakan untuk identifikasi adalah model dengan panjang frame = 275, *overlap* frame = 0.25, k=17 karena pengujian yang digunakan menggunakan data latih yang jumlah data latihnya jauh lebih besar dibandingkan data uji. Hasil pengujian ketika k = 17, panjang frame = 275, dan *overlap* frame = 0.25 ditampilkan pada Tabel 1.

Tabel 1 Hasil pengujian data latih saat k = 17, panjang frame = 275, dan *overlap* frame = 0.25

Uji Data Ke-																								Jenis Tangis				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
'a'	'a'	'a'	'a'	'a'	'a'	'a'	'a'	'a'	'a'	'a'	'a'	'a'	'a'	'n'	'a'	'a'	'n'	'e'	'a'	'a'	'a'	'a'	'a'	'a'	'a'	'a'	'a'	a-eairh
'e'	'e'	'e'	'n'	'e'	'e'	'e'	'e'	'e'	'e'	'h'	'e'	'o'	'e'	'e'	'n'	'e'	'e'	'e'	'e'	'e'	'e'	'e'	'h'	'e'	'e'	'e'	'e'	e-eh
'h'	'h'	'h'	'h'	'h'	'h'	'h'	'h'	'h'	'h'	'e'	'h'	'h'	'h'	'h'	'h'	'h'	'h'	'h'	'h'	'h'	'h'	'h'	'e'	'h'	'h'	'h'	'h'	h-heh
'n'	'n'	'n'	'n'	'n'	'n'	'n'	'n'	'n'	'n'	'n'	'n'	'n'	'h'	'e'	'n'	'n'	'n'	'n'	'h'	'n'	'n'	'a'	'n'	'n'	'n'	'n'	'n'	n-neh
'a'	'o'	'o'	'o'	'o'	'o'	'o'	'n'	'o'	'n'	'o'	'a'	'n'	'o'	'a'	'o'	'e'	'o'	'o'	'a'	'o'	'o'	'o'	'o'	'a'	'h'	'o'	'o'	o-owh

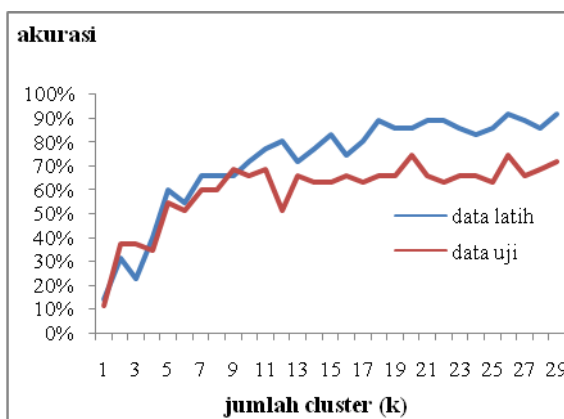
keterangan:  salah identifikasi

Dari beberapa hasil pengujian tersebut dapat diketahui bahwa suara 'heh' merupakan suara yang sering benar dikenali. Sedangkan suara 'owh' merupakan suara yang sering salah dikenali dan biasanya kesalahannya dikenali sebagai suara 'neh' dan 'eairh'. Kesalahan identifikasi ini disebabkan karena variasi data latih 'owh' lebih besar dibandingkan dengan yang lain. Ilustrasi *codebook* 'heh' dan 'owh' pada saat $k=17$, panjang frame = 275, dan *overlap* frame= 0.25 ditampilkan pada Gambar 5.



Gambar 5 Ilustrasi *codebook* 'heh' dan 'owh'

Perbedaan hasil akurasi menggunakan pengujian data latih dan data uji ditampilkan pada Gambar 6. Akurasi tersebut untuk panjang frame = 275, *overlap* frame = 0.25, jumlah *cluster* (k) = 1 s/d 29. Jumlah *cluster* di sini dibuat lebih banyak dengan tujuan supaya dapat menggambarkan perbedaannya.



Gambar 6 Akurasi pengujian data latih dan data uji dengan panjang frame = 275 dan *overlap* frame = 0

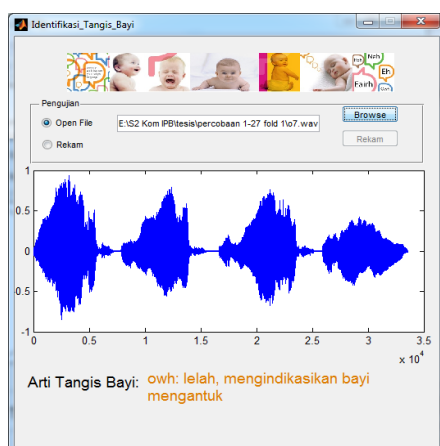
Gambar 6 menunjukkan akurasi menggunakan data latih lebih tinggi dibandingkan data uji. Hal ini dikarenakan pada jarak mahalanobis sudah ada pembagi variasi sehingga akurasi data latih lebih bagus.

Confusion matrix saat $k = 17$, panjang frame = 275, dan *overlap* frame = 0.25 menggunakan jarak mahalanobis ditampilkan pada Tabel 2.

Tabel 2 *Confusion matrix* saat $k = 17$, panjang frame = 275, dan *overlap* frame = 0.25

Jenis Tangis Bayi	Eairh	Eh	Heh	Neh	Owh	Akurasi
Eairh	25	1	0	2	0	89%
Eh	0	23	2	2	1	82%
Heh	0	2	26	0	0	93%
Neh	1	1	2	24	0	86%
Owh	5	1	1	3	18	64%
					Rata-rata	82.80%

Antarmuka identifikasi arti tangis bayi ditampilkan pada Gambar 7. Model codebook adalah model dengan panjang frame = 275, *overlap* frame = 0.25, $k=17$.



Gambar 7 Antarmuka identifikasi arti tangis bayi

SIMPULAN

Software ini digunakan untuk identifikasi jenis tangis bayi versi *Dunstan Baby Language*. Identifikasi ini menggunakan *codebook* untuk pengenalan pola dan MFCC untuk ekstraksi ciri. Penelitian ini memvariasikan panjang frame, *overlap* frame, dan banyaknya *codeword*. Model *codebook* dan MFCC yang menghasilkan akurasi tertinggi yaitu: panjang frame = 275, *overlap* frame = 0.25, $k=17$ dengan akurasi 83%.

DAFTAR PUSTAKA

- Bramer M. 2007. *Principles of Data Mining*. London: Springer.
- Buono A. 2009. Representasi Nilai HOS dan Model MFCC sebagai Ekstraksi Ciri pada Sistem Identifikasi Pembicara di Lingkungan Ber-*noise* Menggunakan HMM. [disertasi]. Depok: Program Studi Ilmu Komputer, Universitas Indonesia.
- Do MN. 1994. Digital Signal Processing Mini-Project: An Automatic Speaker Recognition System. Audio Visual Communication Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland.
- Dunstan P. 2006. Open Up and Discover Your Baby's Language. [Internet]. [diunduh 2012 Maret 19]. Tersedia pada: [http://www.babytaal.nl/media/PDF/ComprehensiveBooklet\(2\).pdf](http://www.babytaal.nl/media/PDF/ComprehensiveBooklet(2).pdf).
- Gomathy M, Meena K, Subramaniam KR. 2012. Gender Clustering and Classification Algorithms in Speech Processing: A Comprehensive Performance Analysis. *International Journal of Computer Applications* 51(20): 9-17.
- Gunawan A. 2011. Dunstan Baby Language Indonesia. [Internet]. [diunduh 2012 September 8]. Tersedia pada: <http://www.mommeworld.com/post/view/49/dunstan-baby-language-indonesia/>.
- Kumar C. Rao PM. 2011. Design of an Automatic Speaker Recognition System using MFCC, Vector Quantization, and LBG Algorithm. *International Journal on Computer Science and Engineering (IJCSSE)* 3(8): 2942-2954.
- Lederman D, Zmora E, Hauschildt S, Stelzig-Eisenhauer A, Wermke K. 2008. Classification of cries of infants with cleft-palate using parallel hidden Markov models. *International Federation for Medical and Biological Engineering* 46: 965-975.
- Lee C, Lien C, Huang R. 2006. Automatic Recognition of Birdsongs Using Mel-frequency Cepstral Coefficients and Vector Quantization. *International MultiConference of Engineers and Computer Scientists*; 2006 June 20-22; Hong Kong.
- Linde Y, Buzo A, Gray RM. 1980. An Algorithm for Vector Quantizer Design. *IEEE Transactions On Communications* 28(1): 84-95.
- Poel M, Ekkel T. 2006. Analyzing Infant Cries Using a Committee of Neural Networks in order to Detect Hypoxia Related Disorder. *International Journal on Artificial Intelligence Tools (IJAIT)* 15(3): 397-410.
- Reyes-Galaviz OF, Reyes-Garcia CA. 2004. A System for the Processing of Infant Cry to Recognize Pathologies in Recently Born Babies with Neural Networks. *International Speech Communication Association*; 2004 September 20-22; Rusia.
- Singh S, Rajan EG. 2011. Vector Quantization Approach for Speaker Recognition using MFCC and Inverted MFCC. *International Journal of Computer Applications* 17(1): 1-7.